EXHIBIT C

Forensic Science International: Genetics xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in

Adele A. Mitchell *, Jeannie Tamariz, Kathleen O'Connell, Nubia Ducasse, Zoran Budimlija, Mechthild Prinz, Theresa Caragine

Department of Forensic Biology, Office of Chief Medical Examiner of The City of New York, 421 E 26th Street, New York, NY 10016, United States

ARTICLE INFO

Article history:
Received 25 May 2012
Received in revised form 8 August 2012
Accepted 9 August 2012

Keywords:
DNA mixture analysis
Likelihood ratio
Allelic drop-out
Allelic drop-in
Low template DNA

ABSTRACT

DNA mixture analysis is a current topic of discussion in the forensics literature. Of particular interest is how to approach mixtures where allelic drop-out and/or drop-in may have occurred. The Office of Chief Medical Examiner (OCME) of The City of New York has developed and validated the Forensic Statistical Tool (FST), a software tool for likelihood ratio analysis of forensic DNA samples, allowing for allelic drop-out and drop-in. FST can be used for single source samples and for mixtures of DNA from two or three contributors, with or without known contributors. Drop-out and drop-in probabilities were estimated empirically through analysis of over 2000 amplifications of more than 700 mixtures and single source samples. Drop-out rates used by FST are a function of the Identifiler® locus, the quantity of template DNA amplified, the number of amplification cycles, the number of contributors to the sample, and the approximate mixture ratio (either unequal or approximately equal). Drop-out rates were estimated separately for heterozygous and homozygous genotypes. Drop-in rates used by FST are a function of number of amplification cycles only.

FST was validated using 454 mock evidence samples generated from DNA mixtures and from items handled by one to four persons. For each sample, likelihood ratios (LRs) were computed for each true contributor and for each profile in a database of over 1200 non-contributors. A wide range of LRs for true contributors was obtained, as true contributors' alleles may be labeled at some or all of the tested loci. However, the LRs were consistent with OCME's qualitative assessments of the results. The second set of data was used to evaluate FST LR results when the test sample in the prosecution hypothesis of the LR is not a contributor to the mixture. With this validation, we demonstrate that LRs generated using FST are consistent with, but more informative than, OCME's qualitative sample assessments and that LRs for non-contributors are appropriately assigned.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The development of increasingly sensitive forensic STR-based DNA testing techniques has expanded the application of DNA typing beyond biological fluids to a wide variety of evidence items such as touched weapons or touched clothes, and forensic laboratories now routinely obtain STR profiles from degraded DNA samples. Increased sensitivity in STR profiling may be accomplished by one of several methods, such as increased PCR cycle numbers [1–4], nested PCR [5,6] and purification of the PCR product [7]. With increased cycle number, full STR profiles can be obtained from 25 to 50 pg of DNA; partial profiles may be obtained from lower quantities of DNA [8,9]. However, stochastic events in early PCR cycles can result in lost alleles (drop-out) and increased sensitivity can produce extraneous alleles (drop-in) [2]. Due to a higher occurrence of allelic drop-out and drop-in with low template or degraded samples, relative to high template, robust samples, the DNA commission of the International Society of Forensic Genetics (ISFG) cautions that standard STR analysis methods should not be directly applied to low template samples [10]. At the Office of Chief Medical Examiner (OCME) of The City of New York, samples containing up to 100 pg of template DNA per amplification are considered low template (LT) DNA and are amplified in triplicate for 31 PCR cycles. Samples containing at least 100 pg of template DNA per amplification are considered high template (HT) DNA and are amplified once or in duplicate for 28 PCR cycles.

The standard statistic calculated when evidentiary and exemplar STR profiles are identical is the random match probability. This can be used for single source evidentiary profiles and for mixtures when individual contributors' profiles can be deduced. Two methods, Combined Probability of Inclusion (CPI), also known as

* Corresponding author.
E-mail address: AAMitchell@ocme.nyc.gov (A.A. Mitchell).

1872-4973/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fsigen.2012.08.007

Please cite this article in press as: A.A. Mitchell, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.08.007

Random Man Not Excluded (RMNE), and likelihood ratio (LR), are commonly used to quantify the statistical weight of mixed DNA profiles when the profiles of individual contributors cannot be deconvoluted. The Scientific Working Group on DNA Analysis Methods (SWGDAM), which is organized by the FBI, published guidelines for interpretation of DNA mixtures in 2010 [11]. The guidelines specify that a statistic should accompany all positive associations between individuals and evidence sample mixtures and that either RMNE or LR may be used. The DNA commission of the ISFG recommends using the LR for such mixtures [10], as more available data are utilized and allelic drop-out and drop-in can be explicitly incorporated in the calculation. That said, RMNE does not require specification of the number of contributors to a mixture, whereas the LR does, and the RMNE calculation is more intuitive; therefore, RMNE can be easier to explain to a jury than the LR.

A likelihood ratio is a ratio of two probabilities. In the numerator is the probability of a set of data conditional on one hypothesis; in the denominator is the probability of the same set of data conditional on a mutually exclusive hypothesis. For forensic DNA applications, the data are the alleles found in the evidence sample, the hypothesis in the numerator is that of the prosecutor (H_p), and the hypothesis in the denominator is that of the defense (H_d). The LR is a measure of the support for the prosecution hypothesis relative to that of the defense. If the LR is greater than one, H_p is better supported by the data than H_d; if the LR is less than one, H_d is better supported by the data than H_p. For single source evidence profiles, the H_p is typically that a particular suspect is the source of the crime scene DNA and H_d is that an unknown, unrelated person is the source of that DNA. For two-person evidence profiles, there are more options for H_p and H_d. First, H_p could be that the crime scene sample represents a mixture of DNA from the suspect and an unknown, unrelated person while H_d could be that the crime scene sample represents a mixture of DNA from two unknown, unrelated people. Alternatively, the prosecutor may assert that the sample represents a mixture of DNA from the suspect and a known person, for example a victim, and the defense may assert that it represents a mixture of DNA from the known person and an unknown, unrelated person or from two unknown, unrelated persons. For three-person evidence profiles, there are even more possibilities, as up to two known contributors may be included in either or both hypotheses. The number of contributors in the two hypotheses need not be the same and a known contributor that is included in either the numerator or the denominator does not need to be included in the other.

OCME has developed and validated the Forensic Statistical Tool (FST), a software tool for LR analysis of low and high template DNA profiles for single source samples and for mixtures of DNA from two or three contributors allowing for drop-out and drop-in. The first step of this development was to empirically determine drop-out rates for each locus, genotype and DNA template-quantity, and drop-in rates for HT-DNA and LT-DNA amplification conditions. These rate estimates were generated from duplicate or triplicate amplifications of over 700 samples, totaling more than 2000 amplifications. These samples include single source samples, as well as two- and three-person deliberate mixtures. Drop-out rates were estimated separately for each Identifiler® locus (Applied Biosystems, Carlsbad, CA), for single source template quantities ranging from 6.25 pg to 500 pg and for two- and three-person mixtures with template quantities ranging 25–500 pg. Separate estimates were obtained for the probability of partial heterozygous drop-out, complete heterozygous drop-out, and complete homozygous drop-out. Drop-out rates were estimated separately for deducible and non-deducible mixtures. Drop-in rates were estimated separately for 28 and 31 cycle amplification.

Empirically estimated drop-out and drop-in rates were incorporated into likelihood ratio frameworks including the appropriate number of contributors (one to three). For mixtures, LRs were formulated with and without assumed contributors ("knowns") in addition to the test profile, usually the suspect profile. For example, for three-person profiles, the LR can be set up as (1) test plus two unknowns versus three unknowns, (2) test plus known plus unknown versus known plus two unknowns, or (3) test plus two knowns versus two knowns plus one unknown.

To date, several other methods and software for LR analysis allowing for drop-out and drop-in have been published. These include True Allele, LoComatioN, the LRMix module within the Forensim package, and LikeLTD. True Allele [12–15] uses Markov chain Monte Carlo (MCMC) to deconvolute a DNA mixture, after which a LR for any set of one to three contributors specified by the user can be computed. In the calculation of the LR, a heuristic penalty for drop-out is applied; for every instance of drop-out that would be required to explain the mixture with a given set of hypothesized contributors, the resulting likelihood ratio is reduced by two orders of magnitude.

LoComatioN [16,17], LRMix [18], likeLTD [19], and FST do not deconvolute DNA mixtures, but simply compute a LR for scenarios specified by the user, allowing for mismatches between contributors' profiles and the DNA alleles labeled in the mixtures. The mismatches are accounted for by incorporating drop-out and drop-in probabilities in the LR calculation. While FST uses empirically determined drop-out and drop-in rates, LoComatioN and Forensim require the user to specify drop-out and drop-in probabilities. Forensim then calculates the LR for a range of drop-out rates and displays the results graphically. LikeLTD finds the drop-out probabilities and mixture ratios that maximize the likelihood under H_p and H_d. One difference between FST and the other three programs is that they adjust for intra- and inter-individual correlation in genotypes [16,20], while FST adjusts only for intra-individual correlation with a correction to the expected population frequencies of homozygous genotypes [11,20]. In addition, LoComatioN and LRMix model allelic drop-in using estimated allele frequencies, whereas FST does not consider the identity of drop-in alleles, simply that drop-in of one allele or of two or more alleles has occurred. In addition, other models [16–18] specifically exclude stutter from the definition of drop-in, whereas FST's drop-in definition includes stutter as well as extraneous peaks that are not in stutter position (see Section 2.2.3).

The FST software was validated by running the program 516 times using 454 mock evidence profiles, including single source samples, deliberate mixtures of DNA from two or three individuals, and samples developed from touched objects handled by one to four known individuals. These samples were composed of various combinations of 85 donors, representing a mix of Caucasian, Asian, African American, and Hispanic ethnicities. FST performance was first evaluated using each true DNA contributor's profile as the test profile and all results were compared to previously generated qualitative conclusions. Other validation steps explored the effect on the FST calculation if the source of the test sample did not contribute to the DNA mixture in question. This was achieved by running FST using each individual in a population database of 1246 non-contributors, collected by OCME and NIST [21] as test profiles. In total, more than 557,000 test runs of the program were performed as part of this validation. We demonstrate that FST is an effective tool for assigning weight of evidence to forensic DNA profiles using likelihood ratios.

2. Methods

2.1. Statistical methods

As a matter of background, in a conventional forensic DNA LR framework, the probability of a crime scene DNA profile (G) is

Please cite this article in press as: A.A. Mitchell, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.08.007

computed conditional on two competing hypotheses: that of the prosecution (H_p) and that of the defense (H_d). The LR is the ratio of these two probabilities: $LR = (Pr(G|H_p)/Pr(G|H_d))$. In a typical simple scenario, the prosecution asserts that the crime scene DNA belongs to the suspect (H_p:S) and the defense asserts that the crime scene DNA belongs to an unknown, unrelated individual (H_d:U). If the crime scene profile matches the suspect profile, then $Pr(G|H_p) = 1.0$ and $Pr(G|H_d) = P$, where P is the estimated population frequency of the crime scene profile (i.e., the random match probability, RMP). Thus, in this scenario, LR = (1/RMP) when the evidence profile matches the suspect profile. When a crime scene sample reflects a mixture, the conventional LR can also be used by specifying the appropriate H_p and H_d. For example, the prosecution may assert that the mixture includes the victim and suspect (H_p:V + S), while the defense may assert the mixture includes the victim and an unknown, unrelated individual (H_d:V + U).

Multiple replicate amplifications of an evidentiary sample can be considered within a LR framework. Let $\mathbf{R} = R_1, R_2, \ldots, R_n$ represent the alleles observed in amplification replicates 1 through n at a single locus in an evidentiary sample. At OCME, normally, $n = 3$ for LT-DNA samples and $n = 1$ or 2 for HT-DNA samples. The replicate data are used to compute $LR = (Pr(\mathbf{R}|H_p)/Pr(\mathbf{R}|H_d))$.

We have incorporated allelic drop-out and drop-in into the LR. A critical step in this process is to consider all possible genotypes for the unknown contributor(s) in the denominator (as well as in the numerator for more complex scenarios). If x distinct alleles are observed at a locus in the evidentiary profile, there are $m = (x(x+1)/2) + x + 1$ values comprising the set of possible genotypes of an unknown contributor. This calculation treats all unobserved alleles as a single 'other' allele. That is, an unknown contributor's genotype at the locus could include any pair wise combination of the observed alleles and the unobserved 'other' allele.

If the prosecution hypothesis is that the DNA in the evidence sample belongs to the suspect (H_p:S) and the defense hypothesis is that the DNA in the evidence sample belongs to an unknown, unrelated person (H_d:U), the likelihood ratio is formulated as:

$$LR = \frac{Pr(\mathbf{R}|S)}{\sum_{j=1}^{m} Pr(\mathbf{R}|U = G_j)Pr(U = G_j)},$$

where S represents the suspect's alleles, U represents the alleles of an unknown contributor, and G_j represents the jth possibility for the genotype of the unknown contributor. $Pr(U = G_j)$ is the expected population frequency of G_j, including a θ correction for intra-individual (but not inter-individual) population substructure, applied to homozygous genotypes, as described in Recommendation 4.1 of the second National Research Council Report [20] and the 2010 SWGDAM mixture interpretation guidelines [11].

If the prosecution hypothesis is that the DNA in the evidence sample belongs to the suspect and the victim (H_p:S + V) and the defense hypothesis is that the DNA in the evidence sample belongs to the victim and an unknown, unrelated person (H_d:V + U), the likelihood ratio is formulated as:

$$LR = \frac{Pr(\mathbf{R}|S, V)}{\sum_{j=1}^{m} Pr(\mathbf{R}|V; U = G_j)Pr(U = G_j)}.$$

If the prosecution hypothesis is that the DNA in the evidence sample belongs to the suspect and an unknown person (H_p:S + U_1) and the defense hypothesis is that the DNA in the evidence sample belongs to two unknown persons, (H_d:U_2 + U_3), the likelihood ratio is formulated as:

$$LR = \frac{\sum_{i=1}^{m} Pr(\mathbf{R}|S; U_1 = G_i)Pr(U_1 = G_i)}{\sum_{i=1}^{m}\sum_{j=1}^{m} Pr(\mathbf{R}|U_2 = G_i, U_3 = G_j)Pr(U_2 = G_i)Pr(U_3 = G_j)}.$$

Table 1
Numerator and denominator options supported by FST. "Known" refers to an elimination profile from an individual who is assumed to be a contributor to the evidence sample. "Comparison" refers to the comparison profile of interest (often the suspect). "Unknown" refers to a randomly selected individual from a population of individuals that are unrelated to the known, comparison, or one another.

Option	Numerator (prosecution hypothesis)	Denominator (defense hypothesis)
1	Comparison	Unknown
2	Comparison + known	Known + unknown
3	Comparison 1 + comparison 2	2 unknowns
4	Comparison + unknown	2 unknowns
5	Comparison + 2 unknowns	3 unknowns
6	Comparison + known + unknown	2 unknowns + known
7	Comparison + 2 known	Unknown + 2 knowns
8	Comparison + 3 known	Unknown + 3 knowns
9	Comparison + 2 knowns + unknown	2 unknowns + 2 knowns
10	Comparison + known + 2 unknowns	3 unknowns + known
11	Comparison + 3 unknowns	4 unknowns

All of the pairs of prosecution and defense hypotheses shown in Table 1 are formulated similarly.

Drop-out and drop-in rates are incorporated into $Pr(R_1, R_2, \ldots, R_n|S)$ and $Pr(R_1, R_2, \ldots, R_n|U = G_j)$. Separate parameters are used for the rate of partial drop-out of heterozygotes, complete drop-out of heterozygotes, and complete drop-out of homozygotes. Drop-out rates were estimated empirically as a function of the locus, the quantity of template DNA, the number of contributors to the sample, and the approximate mixture ratio, i.e., approximately one-to one (not deducible) or not one-to-one (deducible). For single source samples, drop-out rates were estimated for eight template DNA quantities, ranging from 6.25 to 500 pg. For mixtures, drop-out rates were estimated for six template DNA quantities, ranging from 25 to 500 pg. Drop-in rates were estimated separately for LT-DNA (less than or equal to 100 pg per reaction) and HT-DNA (greater than or equal to 100 pg per reaction) amplifications. Note that 100 pg samples were typed under LT-DNA and HT-DNA conditions and two sets of drop-out rates were established for this quantity of template DNA. In total, more than 2000 amplifications were performed for the estimation of drop-out and drop-in rates.

Listed below is a single-locus sample calculation of the LR for a single source evidence profile with H_p: suspect and H_d: unknown. The suspect's alleles at this locus are 11, 14 and the evidence alleles are replicate 1:11; replicate 2:11, 14; replicate 3:11, 14, 15. The drop-out and drop-in that would be required to explain the evidence if the prosecution hypothesis is correct (i.e., the factors in the numerator) are shown in Table 2. Conditional on the evidence sample DNA originating from the suspect, there would be one drop-out from a heterozygous locus and no drop-in in the first replicate, no drop-out from a heterozygous locus and no drop-in in the second replicate, no drop-out from a heterozygous locus and one drop-in in the third replicate. Thus, the numerator in this example is

$$[Pr(D_1)Pr(C_0)] \times [Pr(D_0)Pr(C_0)] \times [Pr(D_0)Pr(C_1)],$$

Table 2
Factors used in numerator (prosecution hypothesis) in example. D_0 and D_1 represent drop-out of zero and one alleles, respectively, from a heterozygous locus; C_0 and C_1 represent drop-in of zero and one alleles, respectively.

Replicate	Alleles labeled	Counts		Factor used	
		Drop-out	Drop-in	Drop-out	Drop-in
1	11	Yes (1)	No	D_1	C_0
2	11, 14	No	No	D_0	C_0
3	11, 14, 15	No	Yes (1)	D_0	C_1

Table 3
Factors used in denominator (defense hypothesis) in example. Variables are defined as in Table 2 plus D_2 which represents drop-out of both alleles from a heterozygous locus, D_{H0} and D_{H1}, which represent no drop-out and complete drop-out, respectively, from a homozygous locus, and C_{2+} which represents drop-in of two or more alleles.

Unknown Person's genotype	Genotype frequency estimate	Replicate 1		Replicate 2		Replicate 3	
		Drop-out	Drop-in	Drop-out	Drop-in	Drop-out	Drop-in
11, 11	P_{11}^2	D_{H0}	C_0	D_{H0}	C_1	D_{H0}	C_{2+}
11, 14	$2\ P_{11}\ P_{14}$	D_1	C_0	D_0	C_0	D_0	C_1
11, 15	$2\ P_{11}\ P_{15}$	D_1	C_0	D_1	C_1	D_0	C_1
11, *w*	$2\ P_{11}\ P_w$	D_1	C_0	D_1	C_1	D_1	C_{2+}
14, 14	P_{14}^2	D_{H1}	C_1	D_{H0}	C_1	D_{H0}	C_{2+}
14, 15	$2\ P_{14}\ P_{15}$	D_2	C_1	D_1	C_1	D_0	C_1
14, *w*	$2\ P_{14}\ P_w$	D_2	C_1	D_1	C_1	D_1	C_{2+}
15, 15	P_{15}^2	D_{H1}	C_1	D_{H1}	C_{2+}	D_{H0}	C_{2+}
15, *w*	$2\ P_{15}\ P_w$	D_2	C_1	D_2	C_{2+}	D_1	C_{2+}
w, *w*	P_w^2	D_{H1}	C_1	D_{H1}	C_{2+}	D_{H1}	C_{2+}

where D_0 and D_1 represent no drop-out and partial drop-out, respectively, from heterozygous loci and C_0 and C_1 represent drop-in of zero and one alleles, respectively.

For the denominator, FST allows the unknown contributor to have a genotype made up of any pair wise combination of the alleles observed in the replicates and any other allele, represented by the letter *w*. In this example, the unknown contributor could have any of the following genotypes: (11, 11), (11, 14), (11, 15), (11, *w*), (14, 14), (14, 15), (14, *w*), (15, 15), (15, *w*), (*w*, *w*), where *w* represents any allele other than 11, 14, or 15. For each possible unknown contributor genotype, FST determines what type of drop-out and drop-in, if any, would be required to explain the evidence profile if the true source of the DNA is an unknown person with that particular genotype. This is analogous to the analysis done using the suspect genotype in the numerator, shown above. The drop-out and drop-in factors for the denominator of this example are shown in Table 3. Using the notation from Table 3 and allowing the unknown contributor to have any of the genotypes listed above, the denominator in this example is shown below. The denominator includes a single term for each possible unknown contributor genotype. Each term includes a factor representing the population frequency estimate of the genotype and drop-out and drop-in probabilities for each replicate. Without incorporating the adjustment to the expected frequency of homozygote genotypes, the denominator is

$$\begin{aligned} &P_{11}^2 \times [Pr(D_{H0})Pr(C_0)] \times [Pr(D_{H0})Pr(C_1)] \times [Pr(D_{H0})Pr(C_{2+})] \\ &\quad + 2P_{11}P_{14} \times [Pr(D_1)Pr(C_0)] \times [Pr(D_0)Pr(C_0)] \times [Pr(D_0)Pr(C_1)] \\ &\quad + 2P_{11}P_{15} \times [Pr(D_1)Pr(C_0)] \times [Pr(D_1)Pr(C_1)] \\ &\quad \times [Pr(D_0)Pr(C_1)] \cdots P_w^2 \times [Pr(D_{H1})Pr(C_1)] \times [Pr(D_{H1})Pr(C_{2+})] \\ &\quad \times [Pr(D_{H1})Pr(C_{2+})] \end{aligned}$$

D_{H0} and D_{H1} represent no drop-out and total drop-out, respectively, from homozygous loci. Because these events are mutually exclusive and represent all possible outcomes, $Pr(D_{H0}) + Pr(D_{H1}) = 1.0$. D_0, D_1, and D_2 represent no drop-out, partial drop-out, and total drop-out, respectively, from heterozygous loci. Again, these events are mutually exclusive and represent all possible outcomes, so $Pr(D_0) + Pr(D_1) + Pr(D_2) = 1.0$. C_0, C_1, and C_{2+} represent drop-in of zero alleles, one allele, and two or more alleles, respectively, in a single amplification at a single locus. These events are also mutually exclusive and represent all possible outcomes, so $Pr(C_0) + Pr(C_1) + Pr(C_{2+}) = 1.0$.

The method described above has been implemented in a C# program with a web interface. The evidence profile, the comparison profile and profiles of known contributors (if applicable) can be entered manually or uploaded. The user selects prosecution and defense hypotheses, specifies whether or not a mixture is deducible, and enters the quantity of template DNA amplified in each reaction (up to three amplifications per evidence profile). For evidence samples with DNA template quantities that fall between those used for drop-out rate estimation, FST interpolates to determine the appropriate rate to use. In order to be conservative, FST uses the drop-out rate estimate minus one standard deviation for each locus, template DNA quantity, number of contributors, and ratio for mixed samples. The final program was tested by performing hand calculations to verify the expected result based on the algorithms explained above and the user set sample characteristics.

FST can perform a comparison between an evidence profile and a single test profile (with or without assumed contributors) or an evidence profile and a database of DNA profiles, such as a quality control database to check for possible sample contamination. We used this capability to determine the distribution of the LR when a non-contributor was included as the test sample. A database of 1246 population samples was used. This database included 546 profiles (115 Caucasian, 125 African American, 151 Hispanic, and 155 Asian contributors) collected locally and 700 profiles (302 Caucasian, 258 African American, and 140 Hispanic contributors) obtained through NIST's STRbase website [21].

2.2. Laboratory methods

2.2.1. DNA sample collection

DNA extracted from buccal swabs given by volunteer donors was used to generate defined DNA mixtures to be used for drop-out rate determination. Extraction, quantification, amplification, separation, and analysis protocols were identical to those used for the mock evidence samples, as described below. To ensure that mixture ratios were accurate, buccal swab extracts were quantified three times in triplicate and the average of the nine values was used for generating mixtures. Extracts were amplified within one week of quantitation or samples were re-quantified.

Mock evidence profiles were developed from mixtures of DNA extracted from post mortem blood or volunteer buccal specimens and from touched items. The test set of mock evidence samples included 98 two-person mixtures from buccal specimens, 102 three-person mixtures from buccal specimens, three two-person mixtures from post mortem blood, 97 two-person mixtures from buccal specimens with either one or both contributors degraded with UV C irradiation, 91 three-person mixtures from post mortem blood, 15 touched items handled by one person, 19 touched items handled by two persons, 39 touched items handled by three persons, and 31 touched items handled by four persons. In total, 350 mixtures and 104 touched items were included in the validation.

Touched items, either cleaned with 10% bleach followed by water and 70% ethanol or not cleaned, were handled by one or

Please cite this article in press as: A.A. Mitchell, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.08.007

consecutively by two, three, or four members of the laboratory as indicated. Touched samples were swabbed with the NYC OCME's swab (patent pending) pre-moistened with 0.01% SDS. Swabbing was performed with a light touch and, if applicable, along the grain of the item. If needed, more than one swab per item or section of an item was used. DNA was extracted from swabs within 1–2 days of collection.

2.2.2. DNA typing

All work was performed using contamination control measures as described by Caragine et al. [9]. Blood and buccal specimens and touched-item swabs were extracted with Chelex, Qiagen M48, or OCME's High Sensitivity Extraction protocol [22] using 0.05% SDS. The digest was purified twice and concentrated with a Microcon® 100 (Millipore, Billerica, MA, USA) pretreated with 1 μg of fish sperm DNA and eluted with 20 μL of irradiated water. Two μL of sample was measured on the Rotor-Gene Q 3000® (Qiagen, Valencia, CA, USA) using an Alu-based real time PCR assay based on the method described by Nicklas and Buel [23], with the exception of the addition of 0.3 μL of 100X SYBR green I (Molecular Probes) and 0.525 mg/mL BSA in a 25 μL reaction volume. With the exception of one set of the blood mixtures, prior to making mixtures, each contributor's DNA extract was measured in triplicate, three times. For one set of the blood mixtures, equivalent sized punches were taken from bloodstain cards and were combined in ratios such as 2:2:1, for example. The "punches" were extracted together and the resultant extract was measured.

Depending on the target quantity, samples were amplified for 28 cycles (≥100 pg, HT-DNA; ID28) or 31 cycles (≤100 pg, LT-DNA; ID31) using the AmpFlSTR® Identifiler® PCR Amplification Kit (Applied Biosystems, Carlsbad, CA). The manufacturer's recommendations were used with the exception of a two minute annealing time, a half-reaction volume, and 2.5 U of AmpliTaq Gold®. ID31 samples were amplified in triplicate and replicates were termed "a", "b", and "c". All other samples were amplified in duplicate.

Amplification products were separated by capillary electrophoresis on an Applied Biosystems (Carlsbad, CA) 3130xl Genetic Analyzer as described previously [9]. Injection parameters were modified based on DNA input and amplification parameters as follows: "ID31 low (L)": 100 pg and 75 pg samples – 1 kV for 22 s, "ID31 normal (N)": 50 pg and 25 pg samples – 3 kV for 20 s, or "ID31 high (H)": 12.5 pg and 6.25 pg samples – 6 kV for 30 s. For samples amplified using Identifiler® 28 cycles (ID28), 3 μL of PCR product were added to HIDI formamide and GeneScan® 500 LIZ® Size Standard (Applied Biosystems, Carlsbad, CA) for a total volume of 30 μL and unless otherwise indicated, injected according to the following parameters: "ID28 high (IR)": samples 200 pg and below – 1 kV for 22 s, or "ID28 normal (I)": samples >200 pg – 5 kV for 20 s.

Data were collected with non-variable binning that collects the highest signal for each color as opposed to variable binning that compensates for low red peak heights. Analysis was performed with Applied Biosystems GeneScan® and Genotyper® or GeneMapper® software with a 75 RFU detection threshold, a 251 baseline window size, a 10% general filter which removes peaks that are less than 10% of the highest peak at a locus, and the OCME's standard locus-specific stutter filters for Identifiler® 28 cycles and for Identifiler® 31 cycles. If multiple injections of a given PCR product were generated for a sample, for each locus the injection or amplification that showed the greatest number of labeled peaks that were not off scale or over saturated was used. For LT-DNA samples, a consensus profile was recorded which contained alleles that repeated in at least two of the three replicates. Additional details on OCME's LT-DNA protocols can be found in Caragine et al. [9].

2.2.3. Drop-out and drop-in rate determination

Drop-out rates were estimated for single source samples with 6.25 pg, 12.5 pg, 25 pg, 50 pg, 100 pg, 150 pg, 250 pg, and 500 pg of template DNA and for two- and three-person mixtures with 25 pg, 50 pg, 100 pg, 150 pg, 250 pg, and 500 pg of template DNA. Mixture ratios of 1:1 (non-deducible) and 4:1 (deducible) were used for two-person mixtures; ratios of 1:1:1 (non-deducible) and 5:1:1 (deducible) were used for three-person samples. Over 700 samples of various combinations of 85 contributors were amplified in duplicate or triplicate and analyzed for the purposes of drop-out and drop-in rate estimation. The 85 contributors represented the diverse population of New York City. For 72% of the samples, the ethnicity of the donor was known, as these donors were laboratory employees. The breakdown was as follows: 20% Asian, 16% black, 54% Caucasian, and 10% Hispanic [24]. The remaining samples were obtained at autopsy and represented a random draw from the population of New York City. According to the 2010 United States census, the population of New York City is 9.8% Asian, 26.6% black, 44.7% Caucasian, and 27% Hispanic [25].

Allele calls for the single source samples and the mixtures were made by the software and artifacts were edited without reference to the known contributors and were then compared to the profiles of the known contributors. Any of the known contributors' alleles that were not present or were present but did not meet the laboratory's limit of detection, 75 RFU, were considered drop-out. Any peaks above 75 RFU that could not be attributed to the known contributors were considered drop-in, whether or not they were in the stutter position of a true allele. Drop-in was modeled this way, as it is never possible to distinguish with certainty whether an extraneous allele in stutter position is actually stutter or not. In addition, for the purposes of the LR, the nature of an extraneous allele is not important.

Drop-out rates were calculated by quantity of template DNA from 25 pg to 500 pg for mixtures and 6.25–500 pg for single source samples. At each locus, the number of opportunities for homozygous and heterozygous drop-out (i.e., the number of homozygous and heterozygous contributors to the samples times the number of amplifications) were counted and used as denominators for the drop-out rates. The number of instances of partial heterozygous drop-out, total heterozygous drop-out, and homozygous drop-out were then tallied and used as the numerators for the drop-out rates. The standard deviation was also calculated and the rates minus one standard deviation were used for the FST program.

Drop-in rates were calculated for each locus and sample type and for 28 cycle and 31 cycle amplification protocols. Drop-in was calculated per locus per replicate. The number of single allele drop-in events and the number of events involving two or more drop-in alleles were counted separately.

2.2.4. Analysis of mock evidence samples

Using OCME's current casework guidelines, mixtures were categorized as deducible or non-deducible. Deducible mixtures can be deconvolved into the full or partial profiles of the major contributors, whereas non-deducible mixtures cannot. The mixture deconvolution was carried out by experienced analysts without any reference to known profiles. The profiles of known contributors were then compared to the mixtures and categorized as one of the following: (a) included/major: the deduced major contributor was consistent with the known contributor and/or all of the alleles of the known contributor were labeled in the mixture; (b) cannot be excluded: most of the known contributor's alleles were labeled in the mixture; (c) no conclusions: several of the known contributor's alleles were not labeled in the mixture but the individual could not be ruled out as a contributor; and (d) excluded: many of the known contributor's alleles were not

Please cite this article in press as: A.A. Mitchell, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.08.007

observed in the mixture. The purpose of the exercise was to perform a comparison to these types of qualitative assessments and demonstrate the effectiveness of probabilistic models.

FST was used to a generate likelihood ratio for each true contributor to each mock evidence profile, based on the apparent number of contributors and deducibility. For two-person profiles, the LR was formulated as suspect + unknown versus two unknowns. For three-person profiles, LR was formulated as suspect + two unknowns versus three unknowns. FST was also tested with a known profile, such as a victim profile, included in the model. For these scenarios, the LR for two-person profiles was formulated as suspect + victim versus victim + unknown and the LR for three-person profiles was formulated as suspect + victim + unknown versus victim + two unknowns.

Furthermore, each mock evidence sample was tested against the database of 1246 non-contributors to determine the distribution of LR expected when a suspect is actually not a contributor to the mixture obtained from the evidence. Therefore, the program was run an additional 1246 times for each evidence sample, using each individual in the population database in turn as the "suspect".

3. Results

For two-person and three-person deducible and non-deducible mixtures, the fraction of alleles called at each STR locus is shown in Fig. 1A–D. Rates are given for deducible two-person samples (A), non-deducible two-person samples (B), deducible three-person samples (C), and non-deducible three-person samples (D). Within each type of mixture, call rates are separated by quantity of template DNA, from 25 pg to 500 pg. Call rates were calculated as $(2A + B - C - 2D - E)/(2A + B)$, where A is the number of

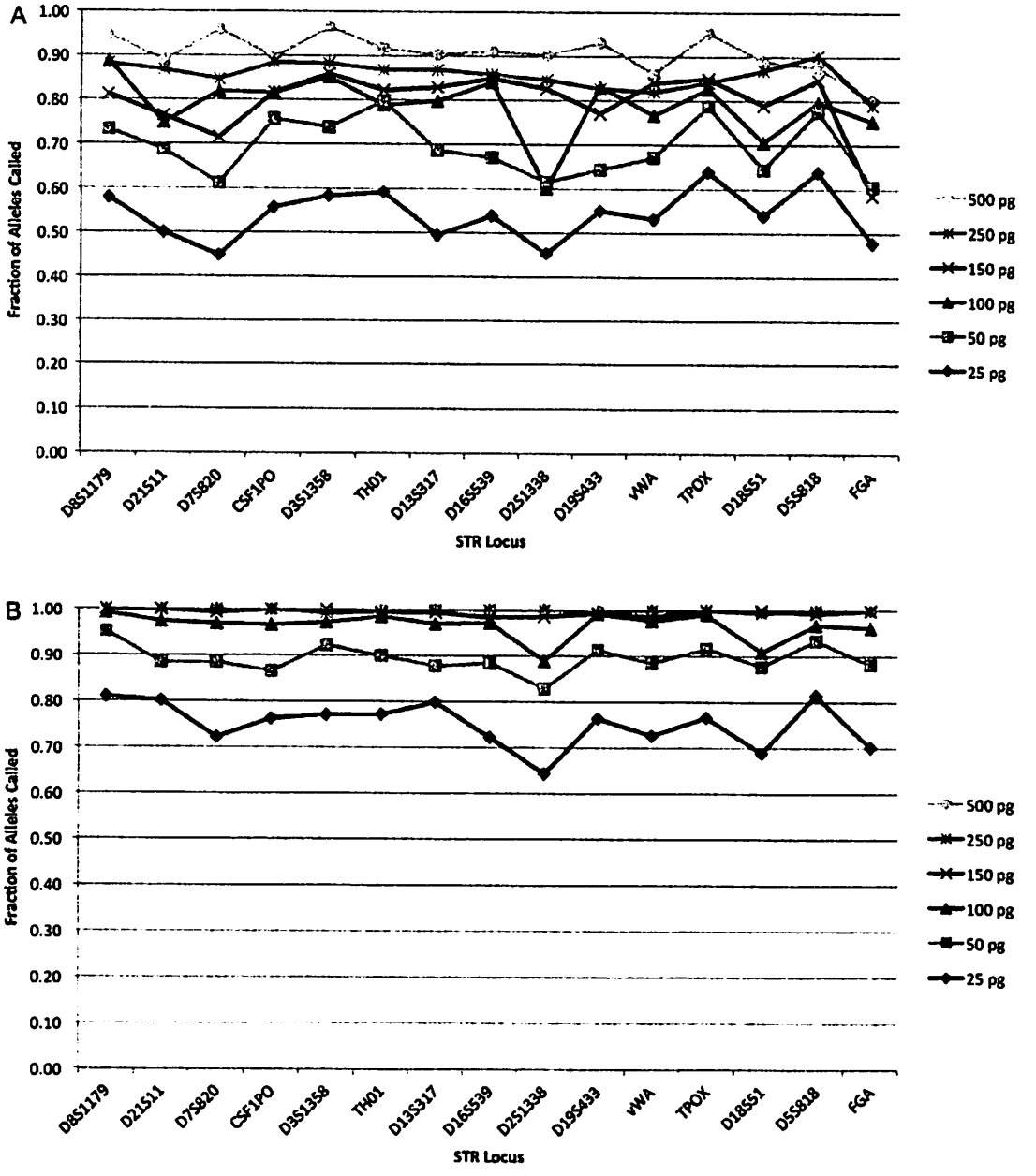

Fig. 1. Fraction of alleles called for minor contributors to two-person deducible mixtures (A), both contributors to two-person non-deducible mixtures (B), minor contributors to three-person deducible mixtures (C), and all contributors to three-person non-deducible mixtures (D).

Please cite this article in press as: A.A. Mitchell, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.08.007

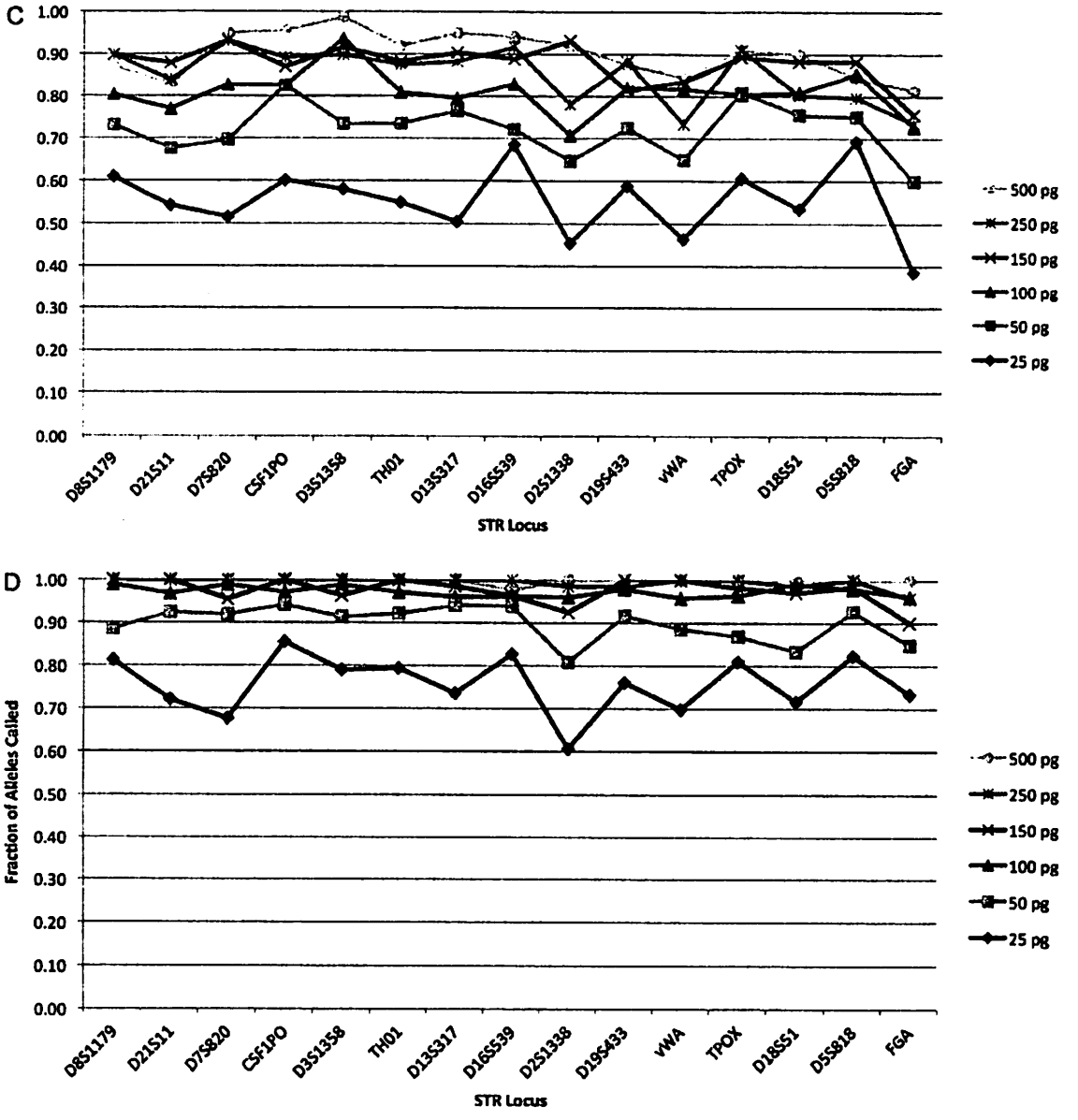

Fig. 1. (*Continued*).

heterozygous loci, B is the number of homozygous loci, C is the number of instances of partial heterozygous drop-out, D is the number of instances of complete heterozygous drop-out, and E is the number of instances of complete homozygous drop-out.

Drop-in rates did not vary significantly by locus, number of contributors, or DNA template quantity. For single source samples, no drop-in alleles were observed with 28 cycle amplification. The average drop-in rate for 31 cycle amplification of single source samples was 0.018 per locus per replicate. For two-person mixtures, the average drop-in rates were 0.009 and 0.041 per locus per replicate for 28 and 31 cycles, respectively. For two-person mixtures, these values were 0.027 and 0.050. Drop-in of two or more alleles at a single locus in a single replicate occurred infrequently, at a rate of approximately 0.005 per locus per replicate. Drop-in rates were not substantially and consistently different for different loci or starting template amounts, but were mainly a function of amplification cycle number. Therefore, only two sets of drop-in rates were incorporated into the FST software and the probability of drop-in of one allele per locus per replicate was set to 0.02 for 28 cycles and 0.035 for 31 cycles amplifications. For both 28 and 31 cycles, 0.005 is used for the probability of drop-in of two or more alleles per locus per replicate. Underestimation of the drop-in rates is likely to be conservative.

The majority of drop-in alleles were found in the minus four stutter position of a true allele. This was true for single source samples and for mixtures, although stutter was more prevalent in mixtures than in single source samples. For single source samples, 60% of drop-in alleles occurred in the minus four position of a true allele. For two-person and three-person mixtures, these values were 77% and 83%. This difference approaches statistical significance ($p = 0.059$). There is a statistically significant difference in minus four drop-in rates for single source samples versus two-person mixtures ($p = 0.009$) and for single source samples versus three-person mixtures ($p = 1.7 \times 10^{-4}$). The fraction of drop-in alleles found in the minus four stutter position was identical (79%) for low template and high template mixtures. Rates were similar across loci.

Please cite this article in press as: A.A. Mitchell, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.08.007

Fig. 2A and B shows the distribution of DNA template quantities amplified for two- (A) and three- (B) person samples in the validation of FST. The maximum length of the whiskers in the box plots is 1.5 times the interquartile range (the distance between the 25th and 75th percentiles). Observations that fall more than 1.5 times the interquartile range away from the 25th or 75th percentile are treated as outliers and indicated by dots beyond the ends of the whiskers. An effort was made to include samples with a wide range of template DNA amounts, with approximately equal numbers of samples in the low template (less than or equal to 100 pg per triplicate amplification) and high template (greater than or equal to 100 pg per duplicate amplification) ranges.

Fig. 3A–D shows the distribution of log likelihood ratios for true contributors with manual calls of "included/major", "cannot be excluded" (CBE), "no conclusions" (NoConc), and "excluded" for two-person deducible mixtures (A), two-person non-deducible mixtures (B), three-person deducible mixtures (C), and three-person non-deducible mixtures (D). Outliers are defined as in Fig. 2. Although there is overlap among the categories, the trend is for decreasing LRs with increasing drop-out.

Log LRs obtained for comparison of non-contributors to mixtures can be found in Fig. 4. Fig. 4A and B shows the distribution of log likelihood ratios for non-contributors to two-person deducible and non-deducible mixtures (A) and to three-person deducible and non-deducible mixtures (B). The number of comparisons to non-contributors was as follows: $N = 166{,}147$ for deducible two-person mixtures, $N = 176{,}158$ for non-deducible two-person mixtures, $N = 93{,}700$ for deducible three-person mixtures, and $N = 121{,}869$ for non-deducible three-person mixtures. Log LRs less than −44 are plotted at −44.

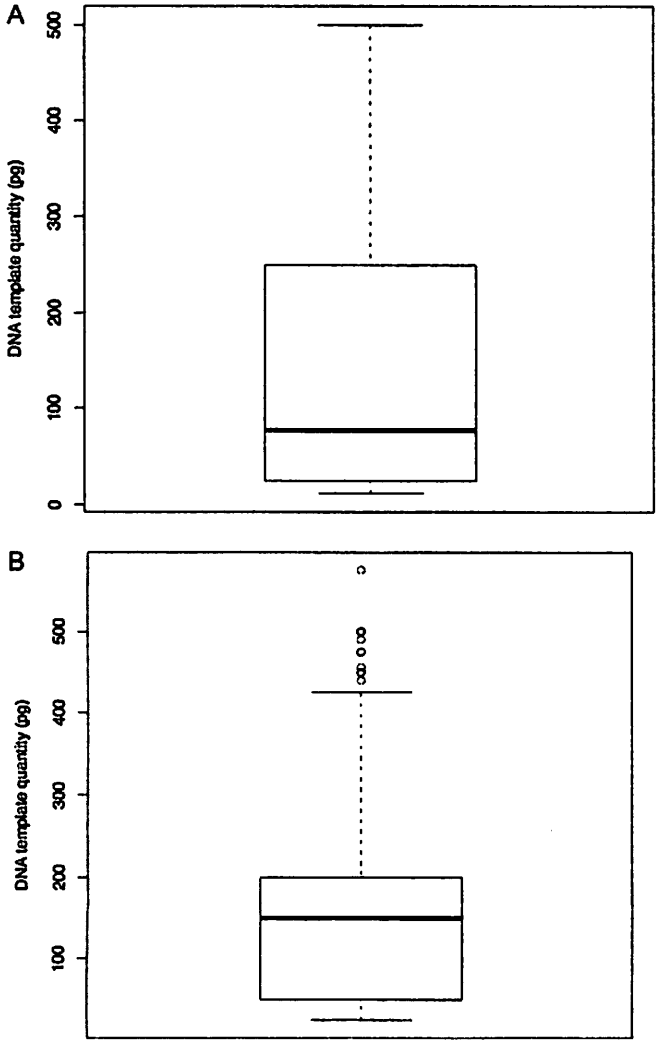

Fig. 2. Distribution of DNA template quantities amplified for two (A) and three (B) person samples. For two-person samples, $N = 272$; for three-person samples, $N = 208$.

Overall, a small number of non-contributors generated a LR greater than 1.0 due to allele sharing, for both the deduced and non-deduced parameters for both two- and three-person mixtures. Although this occurred more frequently with LT-DNA two-person mixtures than with HT-DNA two-person mixtures, there was not a significant difference between these sample types for three-person mixtures. Table 4A and B shows the frequency at which an LR greater than 1, 10, 100, 1000, and 10,000 was obtained for two person (A) and three-person (B) mock evidence samples when non-contributors were used as test samples. Values are shown for deducible and non-deducible mixtures. When non-contributors were tested, no LR's greater than 1000 were observed for two-person mixtures, and no LR's greater than 10,000 were observed for three-person deducible mixtures. Considering all 557,874 comparisons evaluated, the frequency for which a non-contributor generated an LR greater than 100 was 0.0020% and greater than 1000 was 0.00090%.

4. Discussion

4.1. Drop-out and drop-in rates

FST represents one of the first implementations of the likelihood ratio for analysis of forensic DNA mixtures incorporating allelic drop-out and drop-in. Our approach is unique in that drop-out and drop-in rates have been empirically estimated using our laboratory's casework protocols. The drop-out rate estimates employed by FST depend on DNA template quantity, the number of contributors to a sample, their approximate mixture ratio, STR locus, and the genotypes (heterozygote or homozygote) of the hypothesized contributors. Drop-in rates are a function only of the number of amplification cycles used, as drop-in was not substantially different for single source samples versus mixtures or for different STR loci. Our drop-in values (0.005–0.035) are consistent with Balding and Buckleton's [19] recommendation of less than 0.05 to reflect the relative rarity of this phenomenon.

In general, STR loci with longer amplicons demonstrated a trend toward higher drop-out rates than those with shorter amplicons. However, there are a few exceptions. For example, the vWA and TH01 loci have similar amplicon sizes, but the drop-out rates are higher for vWA than for TH01. This may be due to excessive allele sharing at TH01 relative to vWA. For heterozygous loci, a trend for the longer or shorter allele to preferentially drop out was not observed (data not shown). The samples used for the current analysis were pristine buccal swabs. Results may differ if samples displaying more of the complicated characteristics of many evidence samples. Further study of these phenomena is warranted. Consistent with our previously published results [9], drop-out rates increased with decreasing amounts of template DNA.

Although increased allele sharing in mixtures sometimes led to lower drop-out rates for three-person mixtures relative to two-person mixtures, drop-out rates were generally higher with increasing numbers of contributors for a fixed total amount of template DNA. One exception was for common alleles. For example, the 12 allele at the D5S818 locus has frequency 0.25–0.35 in the four populations included in the OCME database. For 25 pg three-person mixtures, the drop-out rate of this allele was 0.069, whereas for 25 pg two-person mixtures it was 0.139. This

Please cite this article in press as: A.A. Mitchell, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.08.002

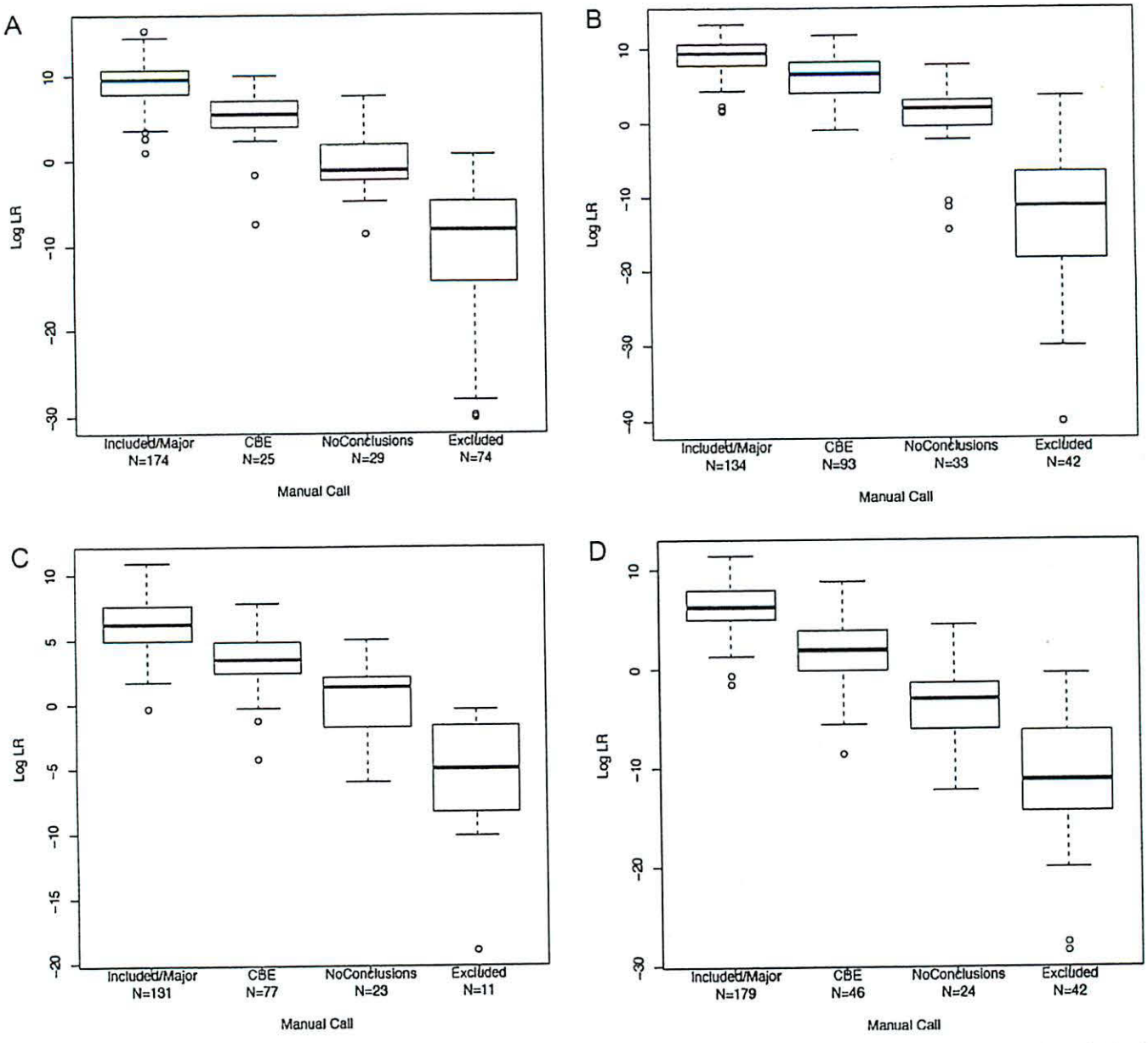

Fig. 3. Distribution of log likelihood ratios for true contributors with manual calls of "included/major", "cannot be excluded" (CBE), "no conclusions" (NoConc), and "excluded" for two-person deducible mixtures (A), two-person non-deducible mixtures (B), three-person deducible mixtures (C), and three-person non-deducible mixtures (D).

difference is likely due to more sharing of the 12 allele with three-person mixtures than with two-person mixtures. For less common alleles at this locus, drop-out rates were higher for three-person than for two-person mixtures.

Estimation of drop-out probabilities is a current topic of discussion in the forensic DNA literature [19,26–30]. Gill et al. [26] introduced a methodology for simulating stochastic variation introduced during each of the steps in STR amplification, including DNA extraction, aliquot sampling, and the PCR process itself. Such a model can be used to describe expected stutter peaks and heterozygote balance and to estimate the probability of drop-out as a function of experimental parameters such as the number of PCR cycles and quantity of template DNA. Gill and Buckleton [27] argue against a peak height threshold below which drop-out is considered, and above which it is not, as this creates a situation in which values just below and just above the threshold are treated very differently. Rather, they advocate for the development of continuous drop-out probability functions. Tvedebrink et al. [28] use such a continuous model with average peak height as a logistic regression predictor of drop-out probability at each locus. Tvedebrink et al. [29] extend the model to allow variable numbers of PCR cycles and to model the STR amplification process in greater detail. The LRMix program [30], on the other hand, does not attempt to model drop-out, but leaves estimation of drop-out rates to the user.

Rather than explicitly modeling the STR amplification process, estimating drop-out probability based on peak heights, and/or allowing users to estimate drop-out rates for a particular sample, we elected to empirically estimate drop-out rates as a function of the total amount of template DNA in a sample, the estimated number of contributors to the sample, their approximate ratio (equal or not equal) and STR locus. In developing and validating our

Please cite this article in press as: A.A. Mitchell, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.08.007

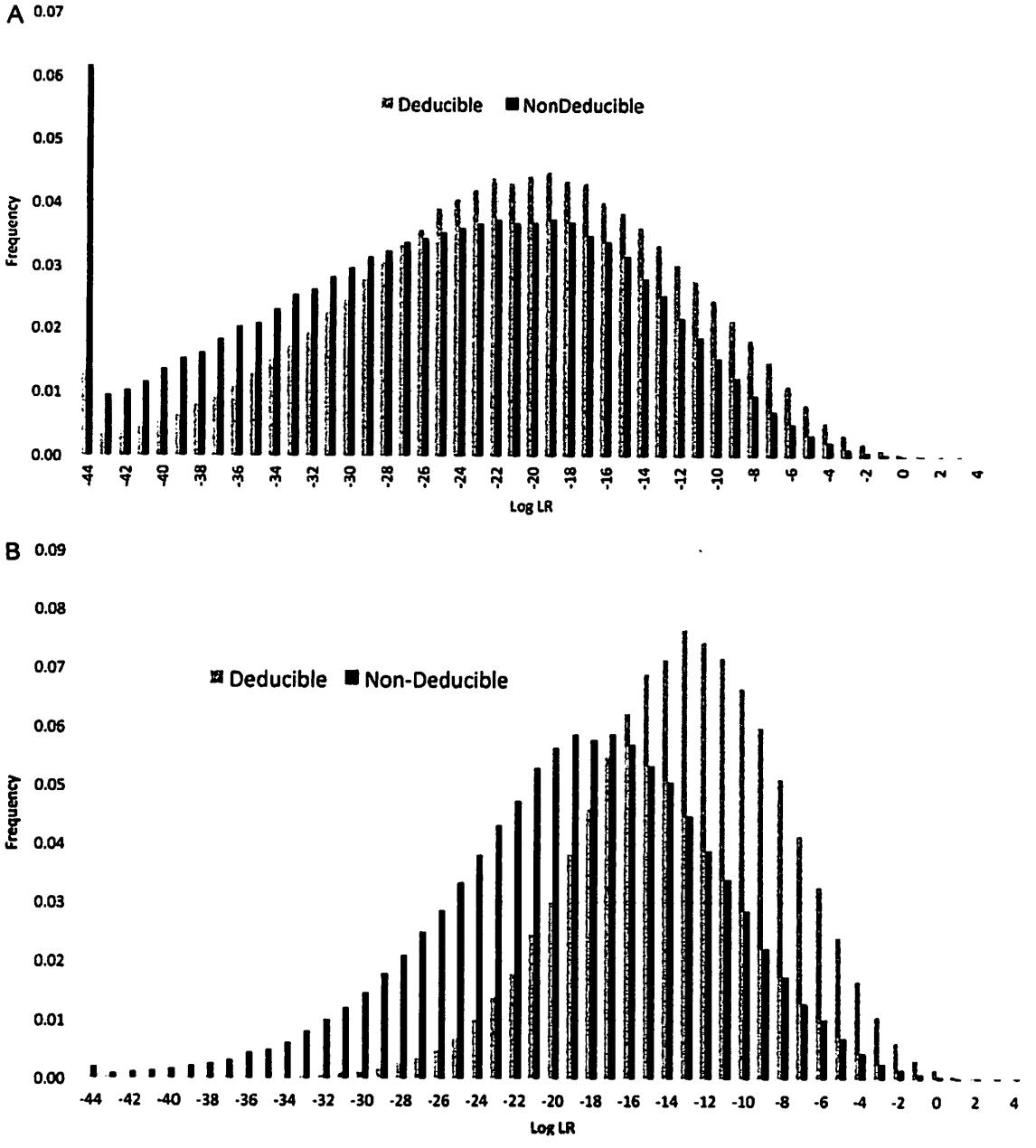

Fig. 4. Distribution of log likelihood ratios for non-contributors to two-person deducible and non-deducible mixtures (A) and to three person deducible and non-deducible mixtures (B). The number of comparisons to non-contributors was as follows: $N = 166{,}147$ for deducible two-person mixtures, $N = 176{,}158$ for non-deducible two-person mixtures, $N = 93{,}700$ for deducible three-person mixtures, and $N = 121{,}869$ for non-deducible three-person mixtures. Log LRs less than −44 are plotted at −44.

high sensitivity testing protocols, we found that peak heights can vary considerably among samples with similar quantities of input DNA and that tall peaks are not necessarily indicative of a low probability of drop-out [9].

OCME's use of a very sensitive and accurate template DNA quantitation system [23] allows reliable quantitation-based estimation of drop-out rates. While there is some variation from instrument to instrument, drop-out rate estimation was performed using data from our laboratory's more sensitive instruments and, thus, the resulting rates may be underestimates for the less sensitive instruments, which is a conservative approach.

Our model for drop-out is continuous, as recommended by Gill and Buckleton [27]. That is, there is no threshold beyond which drop-out is not considered. In order to compute a likelihood ratio for a mixture in which one or more of the hypothesized contributors' alleles are missing, a non-zero probability of drop-out must be used. We use minimum values of 0.05 for drop-out of one allele from heterozygotes and 0.02 for homozygous drop-out and for drop-out of both alleles from heterozygotes. These values were selected after exploration of a range of minimum values. If minimum values are extremely low, a single drop-out event can lead to a very low likelihood ratio, even for true contributors with many repeating alleles. If minimum values are very high, the model is too forgiving of drop-out, leading to low LRs for true contributors without missing alleles and high LRs for non-contributors.

4.2. Validation with mock evidence samples

When FST's performance was tested with mock casework samples that were treated according to OCME casework protocols,

Please cite this article in press as: A.A. Mitchell, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.08.007

Table 4
Frequency at which a likelihood ratio greater than 1, 10, 100, 1000, and 10,000 was obtained for two-person (A) and three-person (B) mock evidence samples when non-contributors were used as comparison samples. Values are shown separately for deducible and non-deducible mixtures.

LR greater than	Observed frequency (1 in x)	
	Deducible	Non-deducible
(A)		
1	5000	10,000
10	16,000	19,000
100	55,000	176,000
1000	>166,000	>176,000
10,000	>166,000	>176,000
(B)		
1	1200	3100
10	4400	7600
100	13,000	40,000
1000	31,000	60,000
10,000	>93,000	121,000

the LR calculations support the qualitative conclusions and help to refine the broad categories of "included/major", "cannot be excluded", "no conclusions", and "excluded". In fact, FST results were often more conservative than manual conclusions, particularly for samples at the top end of HT- and LT-DNA ranges where the FST drop-out rates are low. The LR values were also affected by the presence of common or rare alleles in the mixture.

To determine the distribution of LRs expected when a non-contributor is used as the test profile, the mock evidence profiles were tested against a database of 1,246 population samples. Each non-contributor was treated in turn as the "suspect" with a model that included no known contributors. This exercise is extremely valuable, in that it demonstrates the range of LRs that could be expected when the wrong person is included as a test profile in the model. That is, it gives an idea of what to expect if, for example, a suspect is actually not associated with a piece of tested evidence.

The highest LRs for non-contributors occurred with three-person mixtures, which often contain most of the common alleles at the STR loci. Thus, it would not be unusual to find alleles matching those of a non-contributor at many or all loci in a three-person sample. The highest LR obtained for any non-contributor (4.59×10^4) occurred with an item that had been handled by three-persons without first cleaning the item. This sample contained 575 pg of template DNA per amplification, so it was a high template sample that was amplified in duplicate. The non-contributor that produced the highest LR carried only common alleles that were shared with one or more of the true contributors to the mixture. None of the non-contributor's alleles were missing from the mixture and only two were non-repeating. Thus, this LR appropriately reflects the chance inclusion of this individual's alleles in the mixture.

4.3. Casework considerations

The DNA in evidence samples sometimes shows signs of degradation, which may include decreasing peak heights for loci with longer versus shorter amplicons (i.e., ski slope effect), fewer labeled alleles at longer loci than shorter loci, and STR results of lower quality than expected for the amount of DNA in the sample if the quantification system employed uses shorter amplicons than some of the loci in the STR kit [31,32]. Because degradation is a common phenomenon in evidence samples, a degradation module was incorporated into FST and tested. Using UV light, known samples were degraded and the drop-out rates and patterns of the samples were observed. It was determined that drop-out rates at some loci increased faster than others when samples were subjected to UV radiation, relative to samples that had not been treated. In the FST degradation module, locus-specific drop-out rates were adjusted according to the estimated degree of degradation of a sample, from moderate to severe. Degree of degradation was estimated based on the ratio of peak heights of the longest to the shortest locus in each Identifiler® color. Ultimately, it was determined that, in general, use of the degradation module as programmed resulted in LRs closer to 1.0 for both true contributors and non-contributors. That is, this approach did not increase the overall separation between true contributors and non-contributors (data not shown). This is an area in which further study is warranted, as improvement in quantification of degradation or identification of moderately to severely degraded samples coupled with changes to the degradation model might improve performance.

FST was approved for use with criminal casework samples in New York State by the New York State Forensic Science Commission in December 2010. It is currently in use for mixtures from which major and/or minor contributors cannot be deconvoluted and for comparisons to the minor contributor of mixtures from which a major, but not a minor, contributor can be deconvoluted. Random match probability is still used for single source samples and for deduced profiles.

As explained above all evidence profiles are finalized without consideration of any potential known contributor sample. Samples must be categorized as single source or mixtures and, for mixtures, the number of contributors must be estimated. Characteristics of a mixture can be used to determine whether to treat the mixture as two, three, or four people [24,33–36]. Yet, due to allele sharing amongst related and unrelated individuals, there will always be a level of uncertainty to this determination [33]. Using only the maximum number of alleles observed at any locus to estimate the number of contributors to a mixture will often lead to an underestimate [33,34]. Nevertheless, using the minimum number of contributors typically results in the lowest possible LR, the LR that most favors the defendant [34]. Tools to better estimate the number of contributors are being explored [24,35,36]. FST is currently online for analysis of two- or three-person mixtures. Validation of four-person models is currently in progress.

For mixtures, the analyst must also determine whether or not the profile(s) of major and/or minor contributor(s) can be deduced. If the test profile in a case matches a deduced profile, the random match probability is computed and FST is not used. If the test profile does not match a deduced profile, FST is used with drop-out rates for deducible mixtures. If individual contributors' profiles cannot be determined, FST is used with drop-out rates for non-deducible mixtures. During the FST LR calculation, drop-out is always considered as a possibility, regardless of whether or not any of the potential contributors' alleles are missing. Thus, the model is not selected based on the absence or presence of a suspect or another person's alleles.

Prior to the application of FST to casework samples appropriate prosecution and defense hypotheses must be selected. Within the context of a case, the analyst must determine whether or not to include known contributors in the formulation of the LR. For intimate samples or certain case scenarios, a victim or other individual may be included as a known contributor to a mixture, which will reduce the number of evidence alleles to be explained by the comparison sample. This will generally lead to larger LR values and must be used carefully. Therefore, for certain scenarios, OCME computes and reports the LR with and without including a known contributor, and reports both results.

OCME analysts receive extensive training on presentation of LR results in court. One must be extremely careful when stating any type of results in court and the LR is no different. Great care is taken to avoid transposing the conditional [37] when presenting results.

Please cite this article in press as: A.A. Mitchell, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.08.007

That is, analysts are trained to state that the DNA alleles labeled in the evidence are x times more probable if H_p is true than if H_d is true (or vice versa), rather than stating that H_p is x times more probable than H_d. To help jurors, judges and attorneys understand the magnitude of the results, support for H_p over H_d (or H_d over H_p) is described using the scale of limited support to very strong support suggested by Butler [38] and others.

As an additional interpretation measure, it could be considered to test an evidence mixture against the database of random non-contributors. If the database yields any results larger than those obtained using the test sample from the case, the rate at which this occurred can give an estimate of how often a randomly selected individual could be expected to generate an LR that is at least as large as that obtained using the test sample. For example, if three population samples out of the database of 1246 generated an LR larger than the test sample comparison, we could expect to obtain an LR at least this high for 1 in 415 randomly selected individuals. If the database did not yield any results larger than that of the test sample, the distance between the highest LR for a random population sample and the test sample could be presented. Since testing more contributors would yield a better estimate of the population frequency of the LR results, simulated databases for random non-contributors could also be used. Such simulated DNA profiles as well as Tippet plots were explored by Gill et al. [39] to formulate a different method to test robustness of a weight assessment. Although more work and validation is needed prior to implementation, these general approaches could help put the isolated LR result in perspective.

5. Conclusions

There are two main limitations to the current version of the FST application. First, correlation among genotypes of contributors to mixtures is not considered, which means the calculation is based on unrelated individuals. Each unknown person's genotype is treated as independent from the genotypes of all others in the model. Hypotheses that involve unknown individuals who are related cannot be explicitly modeled and for example, FST cannot accommodate a prosecution hypothesis that includes the victim and the suspect along with a defense hypothesis that includes the victim and the suspect's brother. In this type of situation, the analyst explains that we cannot perform that comparison with the current program, but that the question could probably be resolved with a sample from the brother.

The second limitation of the program is that the drop-out and drop-in rates employed by the program are specific to OCME's protocols, kits and equipment. Application to data generated in another laboratory would require assessment and perhaps adjustment of these rates, as would alteration of OCME's current protocols, kits or equipment. While the program is currently not available to other laboratories, sharing options are being explored and methodology to adapt the parameters for other laboratories or changes to protocols within our laboratory is in progress.

In summary, the FST program performed well over a wide range of DNA template amounts and mixture ratios and types. The results for the touched and purposefully degraded samples demonstrate that the program parameters used are suitable and conservative for degraded samples. The LRs generated when comparing a true contributor with a mixture were generally concordant with manual interpretations. The LR values for true contributors correlated well with the qualitative strength of the positive associations and the few outliers can be explained by the presence of frequent alleles and demonstrates the importance of adding a weight assessment. The addition of a large set of comparisons to random non-contributors identified several chance positive associations mostly to very complex mixtures, as can be expected based on allele sharing, but overall showed very good separation between LR values for contributors and non-contributors. This validation shows that the FST program is a useful tool for calculating a statistical weight for evidence and reference sample comparisons particularly for mixtures that cannot be deconvoluted, or have components that cannot be deconvoluted. The FST software provides a quantitative weight to interpretations that would otherwise be qualitative or where no conclusions could previously be drawn.

Conflict of interest statement

None of the authors has a financial or personal relationship with other people or organizations that could inappropriately influence or bias this paper.

References

[1] I. Findlay, A. Taylor, P. Quirke, R. Frazier, A. Urquhart, DNA fingerprinting from single cells, Nature 389 (1997) 555–556.
[2] P. Gill, J. Whitaker, C. Flaxman, N. Brown, J. Buckleton, An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, Forensic Sci. Int. 112 (2000) 17–40.
[3] J.P. Whitaker, E.A. Cotton, P. Gill, A comparison of the characteristics of profiles produced with the AMPFLSTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis, Forensic Sci. Int. 123 (2001) 215–223.
[4] L. Forster, J. Thomson, S. Kutranov, Direct comparison of post-28-cycle PCR purification and modified capillary electrophoresis methods with the 34-cycle low copy number (LCN) method for analysis of trace forensic DNA samples, Forensic Sci. Int. Genet. 2 (2008) 318–328.
[5] G. Ruano, W. Fenton, K.K. Kidd, Biphasic amplification of very dilute DNA samples via 'booster' PCR, Nucleic Acids Res. 17 (1989) 5407.
[6] P. Taberlet, S. Griffin, B. Goossens, S. Questiau, V. Manceau, N. Escaravage, L.P. Waits, J. Bouvet, Reliable genotyping of samples with very low DNA quantities using PCR, Nucleic Acids Res. 24 (1996) 3189–3194.
[7] P.J. Smith, J. Ballantyne, Simplified low-copy-number DNA analysis by post-PCR purification, J. Forensic Sci. 52 (2007) 820–829.
[8] M. Prinz, L. Schiffner, J.A. Sebestyen, E. Bajda, J. Tamariz, R.C. Shaler, H. Baum, T. Caragine, Maximization of STR DNA typing success for touched objects, Int. Cong. Ser. 1288 (2006) 651–653.
[9] T. Caragine, R. Mikulasovich, J. Tamariz, E. Bajda, J. Sebestyen, H. Baum, M. Prinz, Validation of testing and interpretation protocols for low template DNA samples using AmpFlSTR Identifiler, Croat. Med. J. 50 (2009) 250–267.
[10] P. Gill, C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, B.S. Weir, DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures, Forensic Sci. Int. 160 (2006) 90–101.
[11] Scientific Working Group on DNA Analysis Methods, SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Available from: http://www.fbi.gov/about-us/lab/codis/swgdam-interpretation-guidelines, 2010.
[12] M.W. Perlin, B. Szabady, Linear mixture analysis: a mathematical approach to resolving mixed DNA samples, J. Forensic Sci. 46 (2001) 1372–1378.
[13] M.W. Perlin, J.B. Kadane, R.W. Cotton, Match likelihood ratio for uncertain genotypes, Law Probab. Risk 8 (2009) 289–302.
[14] M.W. Perlin, A. Sinelnikov, An information gap in DNA evidence interpretation, PLoS One 4 (2009) e8327.
[15] M.W. Perlin, M.M. Legler, C.E. Spencer, J.L. Smith, W.P. Allan, J.L. Belrose, B.W. Duceman, Validating TrueAllele(R) DNA mixture interpretation, J. Forensic Sci. 56 (2011) 1430–1447.
[16] J.M. Curran, P. Gill, M.R. Bill, Interpretation of repeat measurement DNA evidence allowing for multiple contributors and population substructure, Forensic Sci. Int. 148 (2005) 47–53.
[17] P. Gill, A. Kirkham, J. Curran, LoComatioN: a software tool for the analysis of low copy number DNA profiles, Forensic Sci. Int. 166 (2007) 128–138.
[18] H. Haned, Forensim: an open-source initiative for the evaluation of statistical methods in forensic genetics, Forensic Sci. Int. Genet. 5 (2011) 265–268.
[19] D.J. Balding, J. Buckleton, Interpreting low template DNA profiles, Forensic Sci. Int. Genet. 4 (2009) 1–10.
[20] National Research Council, The Evaluation of Forensic DNA Evidence, National Academy Press, Washington, DC, 1996.
[21] J.M. Butler, R. Schoske, P.M. Vallone, J.W. Redman, M.C. Kline, Allele frequencies for 15 autosomal STR loci on US Caucasian, African American, and Hispanic populations, J. Forensic Sci. 48 (2003) 1–4.
[22] L.A. Schiffner, E.J. Bajda, M. Prinz, J. Sebestyen, R. Shaler, T.A. Caragine, Optimization of a simple, automatable extraction method to recover sufficient DNA from low copy number DNA samples for generation of short tandem repeat profiles, Croat. Med. J. 46 (2005) 578–586.
[23] J.A. Nicklas, E. Buel, Development of an Alu-based, real-time PCR method for quantitation of human DNA in forensic samples, J. Forensic Sci. 48 (2003) 1–9.

Please cite this article in press as: A.A. Mitchell, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.08.007

Case 1:15-cr-00153-VSB Document 19-3 Filed 12/21/15 Page 14 of 14

[24] J. Perez, A.A. Mitchell, N. Ducasse, J. Tamariz, T. Caragine, Estimating the number of contributors to two-, three-, and four-person mixtures containing DNA in high template and low template amounts, Croat. Med. J. 52 (2011) 314–326.
[25] U.S. Census Bureau, U.S. Census Bureau Delivers New York's 2010 Census Population Totals, Including First Look at Race and Hispanic Origin Data for 2010. Available from: http://2010.census.gov/news/releases/operations/cb11-cn122.html.
[26] P. Gill, J. Curran, K. Elliot, A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci, Nucleic Acids Res. 33 (2005) 632–643.
[27] P. Gill, J. Buckleton, A universal strategy to interpret DNA profiles that does not require a definition of low-copy-number, Forensic Sci. Int. Genet. 4 (2010) 221–227.
[28] T. Tvedebrink, P.S. Eriksen, H.S. Mogensen, N. Morling, Estimating the probability of allelic drop-out of STR alleles in forensic genetics, Forensic Sci. Int. Genet. 3 (2009) 222–226.
[29] T. Tvedebrink, P.S. Eriksen, M. Asplund, H.S. Mogensen, N. Morling, Allelic drop-out probabilities estimated by logistic regression-Further considerations and practical implementation, Forensic Sci. Int. Genet. 6 (2012) 263–267.
[30] H. Haned, T. Egeland, D. Pontier, L. Pene, P. Gill, Estimating drop-out probabilities in forensic DNA samples: a simulation approach to evaluate different models, Forensic Sci. Int. Genet. 5 (2011) 525–531.
[31] M. Graw, H.J. Weisser, S. Lutz, DNA typing of human remains found in damp environments, Forensic Sci. Int. 113 (2000) 91–95.
[32] P.M. Schneider, K. Bender, W.R. Mayr, W. Parson, B. Hoste, R. Decorte, J. Cordonnier, D. Vanek, N. Morling, M. Karjalainen, C. Marie-Paule Carlotti, M. Sabatier, C. Hohoff, H. Schmitter, W. Pflug, R. Wenzel, D. Patzelt, R. Lessig, P. Dobrowolski, G. O'Donnell, L. Garafano, M. Dobosz, P. De Knijff, B. Mevag, R. Pawlowski, L. Gusmao, M. Conceicao Vide, A. Alonso Alonso, O. Garcia Fernandez, P. Sanz Nicolas, A. Kihlgreen, W. Bar, V. Meier, A. Teyssier, R. Coquoz, C. Brandt, U. Germann, P. Gill, J. Hallett, M. Greenhalgh, STR analysis of artificially degraded DNA-results of a collaborative European exercise, Forensic Sci. Int. 139 (2004) 123–134.
[33] J.S. Buckleton, J.M. Curran, P. Gill, Towards understanding the effect of uncertainty in the number of contributors to DNA stains, Forensic Sci. Int. Genet. 1 (2007) 20–28.
[34] H. Haned, L. Pene, J.R. Lobry, A.B. Dufour, D. Pontier, Estimating the number of contributors to forensic DNA mixtures: does maximum likelihood perform better than maximum allele count? J. Forensic Sci. 56 (2011) 23–28.
[35] H. Haned, L. Pène, F. Sauvage, D. Pontier, The predictive value of the maximum likelihood estimator of the number of contributors to a DNA mixture, Forensic Sci. Int. Genet. 5 (2011) 281–284.
[36] D.R. Paoletti, T.E. Doom, C.M. Krane, M.L. Raymer, D.E. Krane, Empirical analysis of the STR profiles resulting from conceptual mixtures, J. Forensic Sci. 50 (2005) 1361–1366.
[37] I.W. Evett, B.S. Weir, Interpreting DNA Evidence. Statistical Genetics for Forensic Scientists, Sinauer Associates, Sunderland, MA, 1998.
[38] J.M. Butler, Forensic DNA Typing, Second Edition: Biology, Technology, and Genetics of STR Markers, Elsevier Academic Press, Burlington, MA, 2005.
[39] P. Gill, J. Curran, C. Neumann, A. Kirkham, T. Clayton, J. Whitaker, J. Lambert, Interpretation of complex DNA profiles using empirical models and a method to measure their robustness, Forensic Sci. Int. Genet. 2 (2008) 91–103.

Please cite this article in press as: A.A. Mitchell, et al., Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in, Forensic Sci. Int. Genet. (2012), http://dx.doi.org/10.1016/j.fsigen.2012.08.007